

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 06-251159

(43)Date of publication of application : 09.09.1994

(51)Int.Cl. G06F 15/70
G01P 13/00
G06F 15/62

(21)Application number : 05-039583

(71)Applicant : NIPPON TELEGR & TELEPH CORP
<NTT>

(22)Date of filing : 01.03.1993

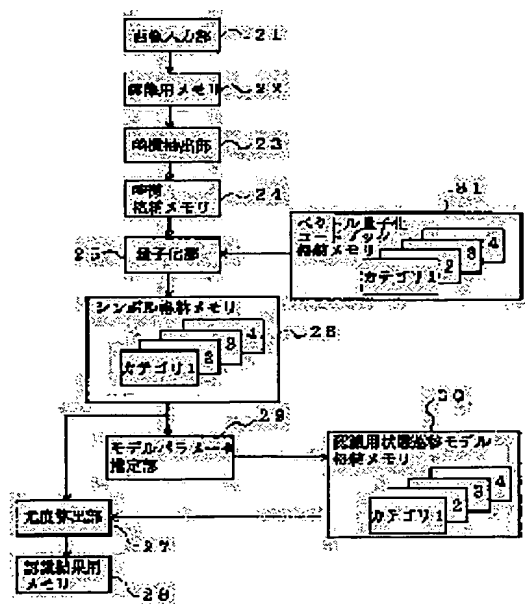
(72)Inventor : YAMATO JUNJI
KURAKAKE MASAHARU
TOMONO AKIRA
ISHII KENICHIRO

(54) OPERATION RECOGNIZING DEVICE

(57)Abstract:

PURPOSE: To improve a success rate of recognition (to reduce a failure) by constituting an operation recognizing system by a dynamic image by learning from an instance.

CONSTITUTION: A code book for quantizing a vector is prepared at every category, and a conversion to a symbol is executed at every category. That is, a feature vector train of a feature storage memory 24 is converted to a symbol train by a quantizing part 25 by referring to a vector quantization code book storage memory 31 corresponding to its category. An output of the quantizing part 25 is stored in a symbol storage memory 26, and it is also different at every used code book. In such a way, a dynamic image is converted to a symbol train, and at the time of its conversion, likelihood is calculated by giving weight corresponding to a distance to the most adjacent representative point in the code book, therefore, by leaning from an instance, an operation of a dynamic body such as a person, etc., in a dynamic image can be recognized stably.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision
of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

* NOTICES *

JPO and NCIP are not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.

2.**** shows the word which can not be translated.

3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] A means to extract characteristic quantity from the input image which photoed the actuation classified into various categories, and to express by the vector, A vector quantization means to change this vector into a symbol train with reference to the code book created beforehand, A means to gain the time series model of operation as a probable state-transition model corresponding to a category by training by study data by using this symbol train as input data, When the image for recognition is inputted, it has a means to calculate respectively the probability for said each model to generate the actuation for recognition. In the recognition equipment of operation out of which the actuation corresponding to a model with the highest likelihood is selected as a recognition result output in the probability for each model to occur The means which prepares the code book of said vector quantization for every recognition category, Recognition equipment of operation characterized [main] by having a means to change a dynamic image into a symbol train using them, and a means to calculate likelihood by giving the weight according to the distance to the recently side representation point in a code book in the case of the conversion.

[Claim 2] It is recognition equipment of operation characterized by being a means by which said feature-extraction means extracts a feature vector in said recognition equipment [according to claim 1] of operation using the mesh description of an image, and an optical flow.

[Translation done.]

*** NOTICES ***

JPO and NCIP are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Industrial Application] This invention relates to the recognition equipment of operation which recognizes the pattern of actuation of animal objects, such as human being from a dynamic image.

[0002]

[Description of the Prior Art] Although, as for the pattern recognition technique for a dynamic image, research of recent years many was done, the artificer was Japanese Patent Application No. 3-205033, in recognition of human being of operation, changed into the symbol train the characteristic quantity vector train acquired from the time series image, and devised the equipment out of which what has the highest likelihood is selected based on the hidden Markov model. Although recognition of the person actuation in a dynamic image was attained by this, recognition success percentage is about 90%.

[0003]

[Problem(s) to be Solved by the Invention] However, by said conventional technique, although the feature vector obtained from a dynamic-image train is distributed over the location where it differs in feature-vector space for every category, distribution may lap between different categories or it may be close. When a bias is in the data for study, the lap of distribution between categories can become large. Thus, when there is a lap of distribution of a feature vector between different categories, assigning the feature vector which belongs to a certain category in the case of the vector quantization in the created code book to the code corresponding to a different category occurs. When data independent of study data were inputted into a recognition system at this time, it had become the big cause of incorrect recognition that it is assigned to the code of the category to which that data belongs, and another similar category, therefore the likelihood of HMM falls.

[0004] Especially, the feature vector changed into the symbol which is not contained in the study pattern of the category concerned in the symbol conversion at the time of recognition occurred, and there was a problem that the likelihood of HMM corresponding to a just category might be lowered. This occupied many of causes of an incorrect recognition example.

[0005] Made in order that this invention may solve said trouble, the purpose of this invention is to offer the technique which can be improved in the success percentage of recognition (reduction of failure).

[0006] Other purposes and new descriptions are clarified by description and the accompanying drawing of this specification at said row of this invention.

[0007]

[Means for Solving the Problem] A means for this invention to extract characteristic quantity from the input image which photoed the actuation classified into various categories, and to express by the vector in order to attain said purpose, A vector quantization means to change this vector into a symbol train with reference to the code book created beforehand, A means to gain the time series model of operation as a probable state-transition model corresponding to a category by training by study data by using this symbol train as input data, When the image for

recognition is inputted, it has a means to calculate respectively the probability for said each model to generate the actuation for recognition. In the recognition equipment of operation out of which the actuation corresponding to a model with the highest likelihood is selected as a recognition result output in the probability for each model to occur. The means which prepares the code book of said vector quantization for every recognition category, It is characterized [main] by having a means to change a dynamic image into a symbol train using them, and a means to calculate likelihood by giving the weight according to the distance to the recently side representation point in a code book in the case of the conversion.

[0008] Said feature-extraction means is characterized by being a means to extract a feature vector using the mesh description of an image, and an optical flow.

[0009]

[Function] Since likelihood is calculated by preparing the code book of vector quantization for every recognition category, changing a dynamic image into a symbol train using them, and giving the weight according to the distance to the recently side representation point in a code book in the case of the conversion according to the above-mentioned means, by study from an example, it is stabilized and actuation of animal objects, such as a person in a dynamic image, can be recognized. Thereby, the success percentage of recognition can be improved (reduction of failure).

[0010] And it differs in that this invention constitutes the code book for vector quantization for every category with a Prior art.

[0011]

[Example] Hereafter, the example of this invention is explained to a detail with reference to a drawing.

[0012] The block block diagram in which drawing 1 shows the outline configuration of one example of the recognition equipment of this invention of operation, and drawing 2 are the block block diagrams showing the functional configuration of this example.

[0013] As for a picture input device and 12, in drawing 1, 11 is [a computer and 13] external memory equipment. drawing 2 -- setting -- 21 -- the image input section and 22 -- the memory for images, and 23 -- the feature-extraction section and 24 -- the description storing memory and 25 -- for the likelihood calculation section and 28, as for the model parameter estimation section and 30, the memory for recognition results and 29 are [the quantization section and 26 / symbol storing memory and 27 / the state-transition model storing memory for recognition and 31] vector quantization code book storing memory.

[0014] There are two phases, study and recognition, in fundamental actuation of this example. At the time of study, parameter estimation of the state-transition model for recognition is performed from the data for study, and it stores in the state-transition model storing memory 30 for recognition for every recognition category. At the time of recognition, the likelihood of the model corresponding to each category stored in the state-transition model storing memory 30 for recognition by study is computed, and maximum likelihood estimation which makes a recognition result the category corresponding to a model with the maximum likelihood is performed. The processing to quantization is the same as that also of the time of recognition at the time of study. Moreover, except the part concerning quantization, it is fundamentally [as Japanese Patent Application No. 3-205033] the same. In accordance with the flow of processing, it explains below.

[0015] First, the dynamic image which contains working human being from the image input sections 21, such as a TV camera, is caught, and it stores in the memory 22 for images.

[0016] Next, two or more characteristic quantity is obtained from a dynamic image by the feature-extraction section 23. Characteristic quantity may be extracted from each frame of a dynamic image, and an image sequence may obtain one characteristic quantity, such as two or more frame lost-motion information respectively corresponding to the sequence of characteristic quantity which case [information etc.] or continues.

[0017] Here, the example of the characteristic quantity to be used is shown below. As an example which obtains one feature vector from one frame, the mesh description shown in

drawing 3 can be considered. That is, the memory 22 for images is first divided into the subblock of $N \times M$ with the number of pixels of $n \times m$, and binary-ization of an image is respectively performed by this subblock. Next, it is the approach of asking for the pulse duty factor of the black pixel in this subblock, and making this the feature vector of a $N \times M$ dimension.

[0018] That is, it is the approach of making a_{ij} the pulse duty factor of the black pixel of a mesh (i, j) , and making a feature vector the vector which put this in order, and $fm = (a_{00}, a_{01}, \dots, a_{ij}, \dots, a_{MN})$.

[0019] After a feature vector is obtained, conversion in the symbol train of a vector train is performed by the quantization section 25, and it is recorded on the memory 26 for symbol storing. This is based on vector quantization. That is, each feature vector is changed into the symbol to which a representation point vector with the nearest distance corresponds based on the list of the representation points for the quantization prepared beforehand. This representation point group is called a code book. By the method of creating a code book The k-mean method (reference 1: X.D.Huang, Y.Ariki and M.A.Jack: "Hidden Markov Models for Speech Recognition", Edingurgh Univ.Press(1990).), The LBG method () [reference 2: Y.Linde,] [A.Buzo] and R.M.Gray: "An Algorithm for Vector Quantization", IEEE There are Trans.Commin., 28, PP.84-85(1980)., etc. In the case of this example, it is necessary to also perform creation of a code book at the time of study of the model for recognition but, and any approach is applicable. Moreover, there are Euclidean distance, Mahalanobis distance in consideration of distribution of each dimension, etc. in the interval scale to be used.

[0020] Then, in this example, like drawing 2, the code book for vector quantization is prepared for every category, and the trouble aforementioned by performing conversion as a symbol for every category is solved. That is, the feature-vector train of the description storing memory 24 is changed into a symbol train by the quantization section 25 with reference to the vector quantization code book storing memory 31 corresponding to the category. Although the output of the quantization section 25 is stored in the symbol storing memory 26, it differs for every code book which also used this.

[0021] Since the input which belongs to the category in order to use the code book for every category is surely assigned to the code of the category, it can abolish the incorrect recognition by quota failure of a code.

[0022] However, about the input of a different category, although the code of that category is assigned, since it is thought that the distance to the representation vector of that code is large, the input belonging to which category can also press down that the likelihood of the data of other categories becomes high by imposing the penalty according to this distance. The detail about this penalty is mentioned later.

[0023] An image sequence is changed into a symbol train by the processing so far. About the actuation so far, the time of study is the same at the time of recognition. About processing after this flowing, the time of recognition is described first.

[0024] At the time of recognition, these feature-vector trains are recorded on the description storing memory 24. And the probability for this feature-vector train to be generated is computed by the likelihood calculation section 27 from each of the model stored in the state-transition model storing memory 30 for recognition for which only the number of categories to recognize was prepared. For the following explanation, the parameter of a model is defined as follows.

[0025] T: The die length of the observed symbol sequence $O (= O_1, O_2, \dots, O_T)$, N: The number of conditions in a model, the number of symbols in L:model, $S = \{s\}$: The set of a condition, The condition (it cannot observe) that s_t is the t -th, $\text{upsilon} = \{\text{upsilon}_1, \text{upsilon}_2, \dots, \text{upsilon}_L\}$: The set of the symbol which can be observed, $A = \{a_{ij} | a_{ij} = \Pr(s_{t+1}=j | s_t=i)\}$: State transition probability, The probability for a_{ij} to change from Condition i to Condition j , $B = \{b_j(O_t) | b_j(O_t) = \Pr(O_t | s_t=j)\}$: A symbol output probability, The probability for $b_j(k)$ to output symbol $\text{upsilon}_{j,k}$ in Condition j , $\pi = \{\pi_i | \pi_i = \Pr(s_1=i)\}$: the probability for a certain model to generate an initial-state probability and a certain observed symbol train With a foward algorithm (reference 1 reference), it is the following, and can make and ask.

[0026] The probability $\Pr(O | \lambda)$ for a certain model $\lambda = [A, B, \pi]$ to output the symbol

sequence O (= O1, O2, ..., OT) is [0027].

[Equation 1]

$$Pr(O|\lambda) = \sum_{i \in S_P} \alpha_T(i) \quad (1)$$

[0028] However, it is here and $\alpha_T(i)$ is [0029].

[Equation 2]

$\alpha_T(i) = Pr(O_1, O_2, \dots, O_t, st=i|\lambda)$ (2)

It comes out and defines and, specifically, is [0030].

[Equation 3]

$$\alpha_t(j) = \left\{ \sum_i \alpha_{t-1}(i) a_{ij} \right\} b_j(O_t) \quad (3)$$

[0031]

[Equation 4]

$\alpha_1 = \pi_i b_i(O_1)$ (4)

It asks by *****.

[0032] However, as a result of preparing a code book for every category in this example as aforementioned, the likelihood of an inaccurate answer category will be esteemed at the time of recognition.

[0033] This is solved multiplying $Pr(O|\lambda)$ by a certain penalty function P_n according to the distance of $f_i(f)$ from the codeword m_j of VQ.

[0034]

[Equation 5]

$Pr'(O|\lambda) = Pr(O|\lambda) \prod_k P_n(f_k)$ (5)

At this example, it is [0035].

[Equation 6]

$$P_n(f) = \exp\left(-\frac{(f - m_j)^2}{\sigma^2}\right) \quad (6)$$

[0036] It carried out. As a result of LBG, since it can be considered that sigma is almost the same, it can be made common to each class and can use the whole average. When much study data are obtained to the dimension of a feature vector, it can be further used for every dimension the whole sigma class, being able to ask.

[0037] In this way, the model with which the called-for likelihood serves as max is chosen as a recognition result, and is stored in the memory 28 for recognition results. That is, in drawing 2, likelihood is calculated for the symbol train changed with the code book for the categories concerned in the likelihood calculation section 27 with the parameter of the state-transition model storing memory 30 for recognition of this category, and this result is outputted. The above is the flow (flow) of the processing at the time of recognition.

[0038] Next, it is ** BE ** about the processing in the case of study flowing. The model parameter estimation section 29 presumes the parameter of a state-transition model which generates the symbol train to the symbol train acquired from the data for study given for every category, and stores it in the state-transition model storing memory 30 for recognition. [two or more] When a certain symbol train O (= O1, O2, ..., OT) is given, a Baum-Welch algorithm is used for this and it is called for (refer to bibliography 1 about a Baum-Welch algorithm). The model for every category for which carried out such and it asked in the model parameter estimation section 29 is stored in the state-transition model storing memory 30 for recognition, and it is used in the case of recognition.

[0039] The example of an experimental result of the flow of the processing stated by this example is explained. In this example, actuation of tennis was used as actuation for recognition.

[0040] In this experiment, a back volley, a backstroke, a forehand volley, a forehand stroke, service, and the example of a smash of operation were used. About four test subjects, ten trial was performed for each [these] category of every, and the data for study and 5 times were

used for 5 times as data for a recognition experiment inside. The mesh description was used as a feature vector. Moreover, the LBG algorithm (reference 2 reference) was used in quantization.

[0041] What has the maximum likelihood in six state-transition models generated by study in six respectively with the application of the data for a recognition experiment was chosen as a recognition result. By changing the combination of a study pattern, ten kinds of recognition systems were constituted and this recognition rate was evaluated.

[0042] The result of an experiment is shown in Table 1. Compared with a conventional method, it can check that the percentage of correct answers has improved.

[0043]

[Table 1]

手法	学習データ			
	A	B	C	D
従来法	100.0	90.6	97.5	91.7
本発明（個別VQ）	100.0	100.0	100.0	98.3

[0044] As mentioned above, although this invention was concretely explained based on the example, this invention is not limited to said example, and can be variously changed in the range which does not deviate from the summary, and things cannot be overemphasized.

[0045]

[Effect of the Invention] As mentioned above, since it can be autonomously adapted for an object or an environment and advanced recognition is attained compared with the conventional technique by constituting the recognition system of operation by the dynamic image by study from an example according to this invention as explained, the success percentage of recognition can be improved (reduction of failure).

[0046] Moreover, since model fitting on an image is not included, robust processing is realizable also to a real image. Therefore, this invention is widely applicable to logging for a right hand side of a request etc. from animations, such as a doubtful actuation monitor in a bank or a store, and a sport.

[Translation done.]

*** NOTICES ***

JPO and NCIP are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] The block block diagram showing the outline configuration of one example of the recognition equipment of this invention of operation,

[Drawing 2] The block block diagram showing the functional configuration of this example,

[Drawing 3] The explanatory view of the mesh description used in the feature-extraction section of this example.

[Description of Notations]

11 [-- The image input section 22 / -- The memory for images, 23 / -- The feature-extraction section, 24 / -- The description storing memory, 25 / -- The quantization section, 26 / -- Symbol storing memory, 27 / -- The likelihood calculation section, 28 / -- The memory for recognition results, 29 / -- The model parameter estimation section, 30 / -- The state-transition model storing memory for recognition, 31 / -- Vector quantization code book storing memory.] -- A picture input device, 12 -- A computer, 13 -- External memory equipment, 21

[Translation done.]

* NOTICES *

JPO and NCIP are not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.

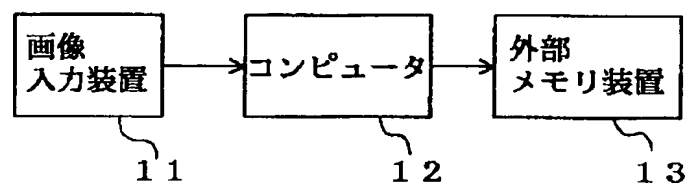
2.*** shows the word which can not be translated.

3.In the drawings, any words are not translated.

DRAWINGS

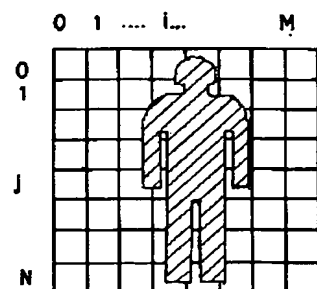
[Drawing 1]

図 1



[Drawing 3]

図 3

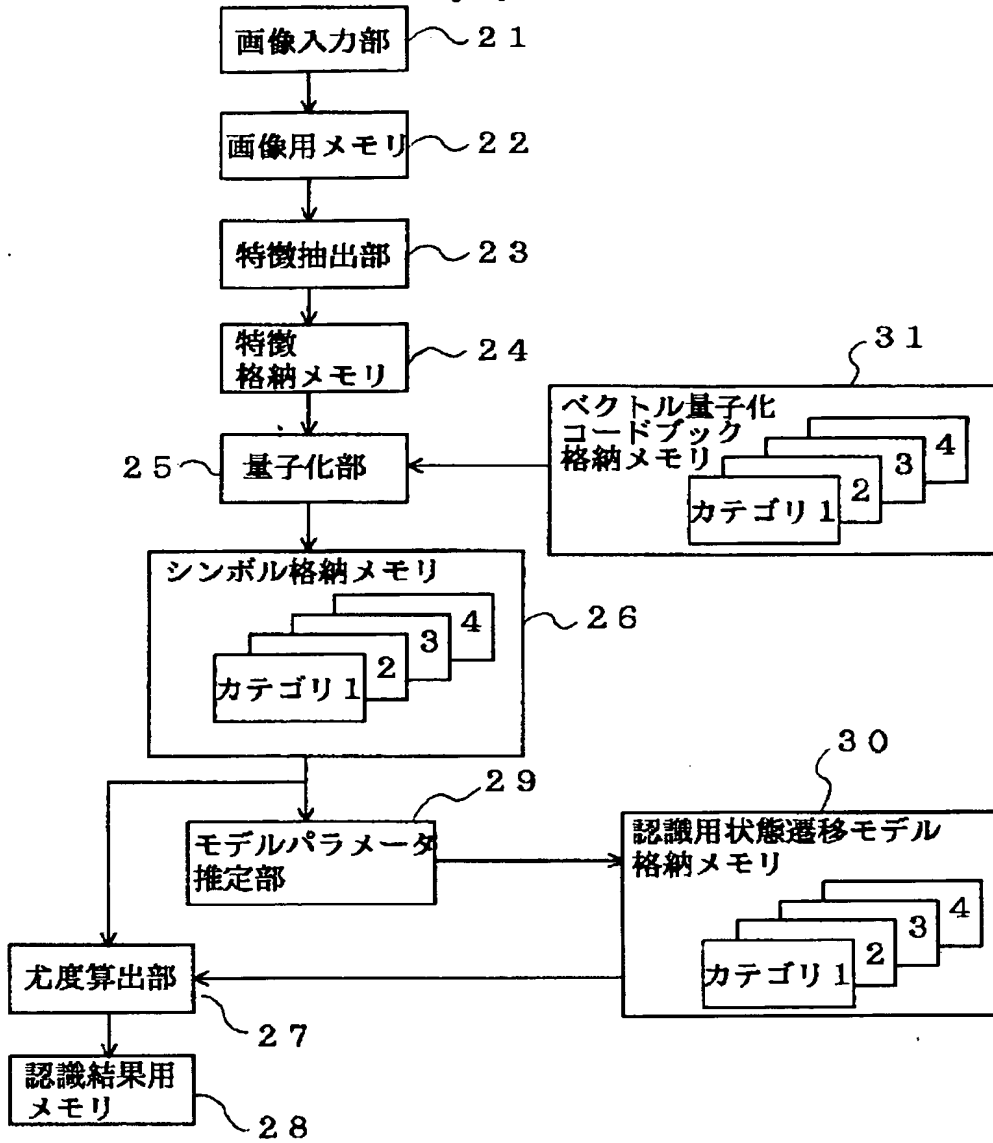


$$f=(a_{00},a_{01},\dots,a_{ij},\dots,a_{MN})$$

$$a_{ij}=\text{number of black mesh}(ij)/M_mN_m$$

[Drawing 2]

図 2



(19)日本国特許庁(JP)

(12) 公開特許公報(A)

(11)特許出願公開番号

特開平6-251159

(43)公開日 平成6年(1994)9月9日

(51)Int.Cl. ⁵	識別記号	庁内整理番号	FI	技術表示箇所
G 0 6 F 15/70	4 6 5 A	8837-5L		
G 0 1 P 13/00	A			
G 0 6 F 15/62	3 8 0	9287-5L		

審査請求 未請求 請求項の数 2 OL (全 6 頁)

(21)出願番号 特願平5-39583

(22)出願日 平成5年(1993)3月1日

(71)出願人 000004226

日本電信電話株式会社
東京都千代田区内幸町一丁目1番6号

(72)発明者 大和 淳司

東京都千代田区内幸町一丁目1番6号 日
本電信電話株式会社内

(72)発明者 倉掛 正治

東京都千代田区内幸町一丁目1番6号 日
本電信電話株式会社内

(72)発明者 伴野 明

東京都千代田区内幸町一丁目1番6号 日
本電信電話株式会社内

(74)代理人 弁理士 秋田 収喜

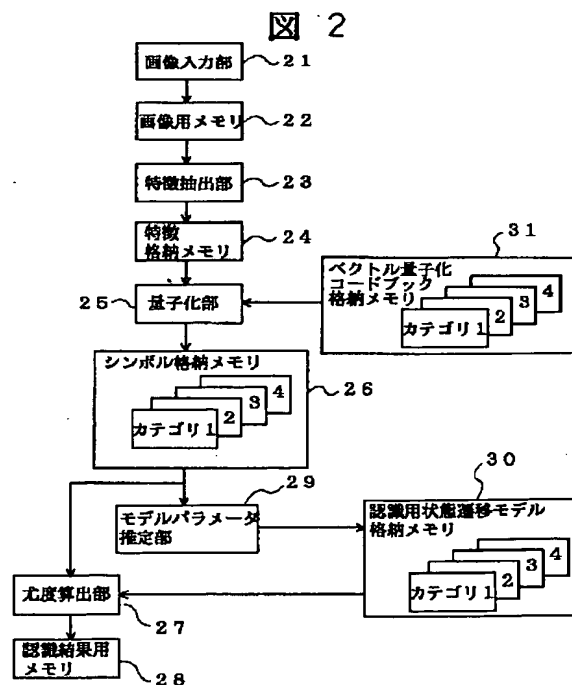
最終頁に続く

(54)【発明の名称】 動作認識装置

(57)【要約】

【目的】 認識の成功率を向上(失敗の低減)する。

【構成】 各種カテゴリに分類される動作を撮影した入力画像から特徴量を抽出し、ベクトルで表現し、該ベクトルを予め作成したコードブックを参照してシンボル列に変換し、該シンボル列を入力データとしてカテゴリに対応する確率的状態遷移モデルとしての動作の時系列モデルを学習データによるトレーニングによって獲得し、認識対象画像が入力された際、前記各モデルが認識対象動作を生成する確率を各々計算し、各モデルが生起する確率の中で最も尤度の高いモデルに対応する動作を認識結果出力として選び出す動作認識装置において、前記ベクトル量子化のコードブックを認識カテゴリ毎に用意しておく手段と、それらを使用して動画像をシンボル列に変換する手段と、その変換の際にコードブック中の最近傍代表点までの距離に応じた重みをつけて尤度を計算する手段を備える。



【特許請求の範囲】

【請求項 1】 各種カテゴリに分類される動作を撮影した入力画像から特徴量を抽出し、ベクトルで表現する手段と、該ベクトルを予め作成したコードブックを参照してシンボル列に変換するベクトル量子化手段と、該シンボル列を入力データとしてカテゴリに対応する確率的状態遷移モデルとしての動作の時系列モデルを学習データによるトレーニングによって獲得する手段と、認識対象画像が入力された際、前記各モデルが認識対象動作を生成する確率を各々計算する手段とを有し、各モデルが生

起する確率の中で最も尤度の高いモデルに対応する動作を認識結果出力として選び出す動作認識装置において、前記ベクトル量子化のコードブックを認識カテゴリ毎に用意しておく手段と、それらを使用して動画像をシンボル列に変換する手段と、その変換の際にコードブック中の最近傍代表点までの距離に応じた重みをつけて尤度を計算する手段を備えることを最も主要な特徴とする動作認識装置。

【請求項 2】 前記請求項 1 に記載の動作認識装置において、前記特徴抽出手段は、画像のメッシュ特徴、オプティカルフローを用いて、特徴ベクトルを抽出する手段であることを特徴とする動作認識装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、動画像からの人間等の動物体の動作のパタンの認識を行う動作認識装置に関するものである。

【0002】

【従来の技術】 動画像を対象とするパターン認識技術は、近年多くの研究が行われているが、発明者は、特願平 3-205033 で、人間の動作認識において、時系列画像から得た特徴量ベクトル列をシンボル列に変換し、隠れマルコフモデルに基づき、最も尤度の高いものを選び出す装置を考案した。これにより動画像中の人物動作の認識が可能となったが、認識成功率が 90% 程度である。

【0003】

【発明が解決しようとする課題】 しかしながら、前記従来の手法では、動画像列から得られる特徴ベクトルは、カテゴリ毎に特徴ベクトル空間中の異なる位置に分布しているが、異なるカテゴリ間で分布が重なったり、近接していることもある。学習用データに偏りがある場合には、カテゴリ間の分布の重なりが大きくなることもあり得る。このように、異なるカテゴリ間で、特徴ベクトルの分布の重なりがある場合には、作成されたコードブックでのベクトル量子化の際に、あるカテゴリに属する特徴ベクトルを、異なるカテゴリに対応するコードに割り当てるが発生する。このとき、学習データと独立なデータを認識系に入力したとき、そのデータが属するカテゴリと類似する別のカテゴリのコードに割り当てら

れ、そのために HMM の尤度が低下することが誤認識の大きな原因となっていた。

【0004】 特に、認識時のシンボル変換において、当該カテゴリの学習パターンに含まれないシンボルに変換される特徴ベクトルが発生し、正当カテゴリに対応する HMM の尤度を下げることがあるという問題があった。これが誤認識事例の原因の多くを占めていた。

【0005】 本発明は、前記問題点を解決するためになされたものであり、本発明の目的は、認識の成功率を向上（失敗の低減）することが可能な技術を提供することにある。

【0006】 本発明の前記ならびにその他の目的及び新規な特徴は、本明細書の記述及び添付図面によって明らかにする。

【0007】

【課題を解決するための手段】 前記目的を達成するために、本発明は、各種カテゴリに分類される動作を撮影した入力画像から特徴量を抽出し、ベクトルで表現する手段と、該ベクトルを予め作成したコードブックを参照してシンボル列に変換するベクトル量子化手段と、該シンボル列を入力データとしてカテゴリに対応する確率的状態遷移モデルとしての動作の時系列モデルを学習データによるトレーニングによって獲得する手段と、認識対象画像が入力された際、前記各モデルが認識対象動作を生成する確率を各々計算する手段とを有し、各モデルが生起する確率の中で最も尤度の高いモデルに対応する動作を認識結果出力として選び出す動作認識装置において、前記ベクトル量子化のコードブックを認識カテゴリ毎に用意しておく手段と、それらを使用して動画像をシンボル列に変換する手段と、その変換の際にコードブック中の最近傍代表点までの距離に応じた重みをつけて尤度を計算する手段を備えることを最も主要な特徴とする。

【0008】 前記特徴抽出手段は、画像のメッシュ特徴、オプティカルフローを用いて、特徴ベクトルを抽出する手段であることを特徴とする。

【0009】

【作用】 前述の手段によれば、ベクトル量子化のコードブックを認識カテゴリ毎に用意しておき、それらを使用して動画像をシンボル列に変換し、その変換の際にコードブック中の最近傍代表点までの距離に応じた重みをつけて尤度を計算するので、事例からの学習によって、動画像中の人物などの動物体の動作を安定して認識することができる。これにより、認識の成功率を向上（失敗の低減）することができる。

【0010】 そして、本発明は、従来の技術とは、ベクトル量子化のためのコードブックをカテゴリ毎に構成する点が異なる。

【0011】

【実施例】 以下、本発明の実施例を図面を参照して詳細に説明する。

【0012】図1は、本発明の動作認識装置の一実施例の概略構成を示すブロック構成図、図2は、本実施例の機能構成を示すブロック構成図である。

【0013】図1において、11は画像入力装置、12はコンピュータ、13は外部メモリ装置である。図2において、21は画像入力部、22は画像用メモリ、23は特徴抽出部、24は特徴格納メモリ、25は量子化部、26はシンボル格納メモリ、27は尤度算出部、28は認識結果用メモリ、29はモデルパラメータ推定部、30は認識用状態遷移モデル格納メモリ、31はベ

クトル量子化コードブック格納メモリである。

【0014】本実施例の基本的動作には、学習と認識の2つの段階がある。学習時には学習用のデータから認識用状態遷移モデルのパラメータ推定を行い認識カテゴリ毎に認識用状態遷移モデル格納メモリ30に格納する。認識時には学習によって認識用状態遷移モデル格納メモリ30に格納された、各カテゴリに対応するモデルの尤度を算出し、最大の尤度を持つモデルに対応するカテゴリを認識結果とする最尤推定を行う。量子化までの処理は、学習時、認識時とも同一である。また、量子化に係る部分以外については特願平3-205033と基本的に同じである。以下処理の流れに沿って説明する。

【0015】まず、TVカメラ等の画像入力部21から動作中の人間を含む動画像をとらえ、画像用メモリ22に格納する。

【0016】次に、特徴抽出部23により、動画像から、複数の特徴量を得る。特徴量は、動画像の各フレームから抽出し、画像系列が特徴量の系列に各々対応する場合、または連続する複数のフレームから動き情報など一つの特徴量を得る場合がある。

【0017】ここで、使用する特徴量の例を以下に示す。1つのフレームから1つの特徴ベクトルを得る例として、図3に示すメッシュ特徴が考えられる。すなわち、まず、画像用メモリ22を $n \times m$ の画素数を持つ $N \times M$ のサブブロックに分割し、各々このサブブロックで画像の2値化を行う。次に、このサブブロック内の黒画素の占有率をもとめ、これを $N \times M$ 次元の特徴ベクトルとする方法である。

【0018】すなわち、 a_{ij} をメッシュ (i, j) の黒画素の占有率とし、これを並べたベクトル、 $f_s = (a_{00}, a_{01}, \dots, a_{ij}, \dots, a_{MN})$ を特徴ベクトルとする方法である。

【0019】特徴ベクトルが得られた後、量子化部25によってベクトル列のシンボル列への変換が行われ、シンボル格納用メモリ26に記録される。これはベクトル量子化による。すなわち、各特徴ベクトルはあらかじめ用意された量子化のための代表点の一覧に基づき、それらの内で最も距離の近い代表点ベクトルの対応するシンボルに変換される。この代表点群をコードブックと呼ぶ。コードブックの作成法には、 k -mean法（文献

1: X. D. Huang, Y. Ariki and M. A. Jack: "Hidden Markov Models for Speech Recognition", Edinburgh Univ. Press (1990).)、LBG法（文献2: Y. Linde, A. Buzo and R. M. Gray: "An Algorithm for Vector Quantization", IEEE Trans. Commun., 28, PP. 84-85 (1980).)などがある。本実施例の場合、認識用モデルの学習時にコードブックの作成も行う必要があるが、いずれの方法も適用可能である。また、使用する距離尺度には、ユークリッド距離、各次元の分散を考慮したマハラノビス距離などがある。

【0020】そこで、本実施例においては、図2のように、各カテゴリ毎にベクトル量子化のためのコードブックを用意し、それぞれのカテゴリ毎にシンボルへの変換を行うことで前記の問題点は解決する。すなわち、特徴格納メモリ24の特徴ベクトル列は、そのカテゴリに対応したベクトル量子化コードブック格納メモリ31を参照して量子化部25によりシンボル列に変換される。量子化部25の出力はシンボル格納メモリ26に蓄えられるが、これも使用したコードブック毎に異なるものである。

【0021】カテゴリ毎のコードブックを用いるため、そのカテゴリに属する入力には必ずそのカテゴリのコードに割り当てられるため、コードの割り当て失敗による誤認識をなくすることができる。

【0022】ただし、どのカテゴリに属する入力もそのカテゴリのコードが割り当てられるが、異なるカテゴリの入力については、そのコードの代表ベクトルまでの距離が大きいと考えられるので、この距離に応じたペナルティを課すことで、他のカテゴリのデータの尤度が高くなることを押えることができる。このペナルティに関する詳細は後述する。

【0023】ここまでの処理によって、画像系列がシンボル列に変換される。ここまでの動作については、認識時、学習時とも同一である。これ以降の処理の流れについて、まず、認識時について述べる。

【0024】認識時には、これらの特徴ベクトル列は、特徴格納メモリ24に記録される。そして、認識するカテゴリ数だけ用意された認識用状態遷移モデル格納メモリ30に格納されたモデルの各々から、この特徴ベクトル列が生成される確率を尤度算出部27によって算出する。以下の説明のために、モデルのパラメータを次のように定める。

【0025】 T : 観測されたシンボル系列 $O = (O_1, O_2, \dots, O_T)$ の長さ、 N : モデル中の状態数、 L : モデル中のシンボル数、 $S = \{s\}$: 状態の集合、 s_t は t 番目の状態（観測できない）、 $v = \{v_1, v_2, \dots, v_L\}$: 観測可能なシンボルの集合、 $A = \{a_{ij} \mid a_{ij} = \Pr(s_{t+1} = j \mid s_t = i)\}$: 状態遷移確率、 a_{ij} は状態 i から状態 j へ遷移する確率、 $B = \{b_j \mid O_t$

$\{b_j(O_t) = \Pr(O_t | s_t = j)\}$: シンボル出力確率、 $b_j(k)$ は状態 j においてシンボル v_k を出力する確率、

$\pi = \{\pi_i | \pi_i = \Pr(s_1 = i)\}$: 初期状態確率、観測したあるシンボル列を、あるモデルが発生する確率は、forward アルゴリズム (文献1 参照) によって以下の

$$\Pr(O|\lambda) = \sum_{i \in S_F} \alpha_T(i)$$

【0028】ただし、ここで $\alpha_T(i)$ は、

【0029】

$$\alpha_T(i) \equiv \Pr(O_1, O_2, \dots, O_T, s_T = i | \lambda) \quad (2)$$

で定義され、具体的には、

【0030】

$$\alpha_t(j) = \left\{ \sum_i \alpha_{t-1}(i) a_{ij} \right\} b_j(O_t) \quad (3)$$

【0031】

$$\alpha_1 = \pi_i b_i(O_1)$$

の漸化式で求められる。

【0032】ただし、前記のとおり、本実施例において、コードブックをカテゴリ毎に用意した結果、認識時には不正答カテゴリの尤度が高く評価されることになる。

本実施例では、

【0035】

【数6】

$$P_n(f) = \exp\left(\frac{(\sigma - m_j)^2}{\sigma^2}\right) \quad (6)$$

【0036】とした。LBGの結果 σ はほぼ同じとみなせるので、各クラス共通とし、全体の平均を使用することができる。特徴ベクトルの次元に対して学習データが多数得られる場合には、 σ クラス毎、さらに次元毎に求めて使用することができる。

【0037】こうして求められた尤度が最大となるモデルが、認識結果として選択され認識結果用メモリ28に蓄えられる。すなわち、図2においては、尤度算出部27において当該カテゴリ用のコードブックにて変換されたシンボル列を該カテゴリの認識用状態遷移モデル格納メモリ30のパラメータにて尤度を計算し、この結果を出力する。以上が認識時の処理の流れ(フロー)である。

【0038】次に、学習の際の処理の流れについて述べる。モデルパラメータ推定部29は、各カテゴリ毎に複数与えられた学習用データから得られたシンボル列に対して、そのシンボル列を発生するような状態遷移モデルのパラメータを推定し、認識用状態遷移モデル格納メモリ30に蓄える。これは、あるシンボル列 $O = (O_1, O_2, \dots, O_T)$ が与えられたときに、Baum-Welchアルゴリズムを用いて求められる (Baum-Welchアルゴリズム

ようにして求めることができる。

【0026】あるモデル $\lambda = \{A, B, \pi\}$ がシンボル系列 $O = (O_1, O_2, \dots, O_T)$ を出力する確率 $\Pr(O | \lambda)$ は、

【0027】

【数1】

(1)

【数2】

【数3】

【数4】

(4)

20 【0033】VQのコードワード m_j から f_i の距離に応じたなんらかのペナルティ関数 $P_n(f)$ を $\Pr(O | \lambda)$ にかけてこれを解決する。

【0034】

【数5】

$$\Pr'(O | \lambda_i) = \Pr(O | \lambda_i) \prod_k P_n(f_k) \quad (5)$$

ムについては参考文献1 参照のこと)。このようにしてモデルパラメータ推定部29にて求めた各カテゴリごとのモデルを認識用状態遷移モデル格納メモリ30に格納し、認識の際に使用する。

30 【0039】本実施例で述べた処理の流れの実験結果例を説明する。本実施例では、認識対象の動作として、テニスの動作を使用した。

【0040】本実験では、バックボレー、バックストローク、フォアボレー、フォアストローク、サービス、スマッシュの動作例を用いた。4人の被験者について、これら各カテゴリ毎に10回の試行を行い、内5回を学習用データ、5回を認識実験用データとして使用した。特徴ベクトルとしてはメッシュ特徴を使用した。また、量子化においては、LBGアルゴリズム (文献2 参照) を使用した。

【0041】学習によって生成された6つの状態遷移モデルに、認識実験用のデータを適用して各々、6つのなかで最大の尤度をもつものを認識結果として選択した。学習パタンの組合せを変えることにより10通りの認識系を構成し、この認識率を評価した。

【0042】実験の結果を表1に示す。従来法に比べ、正答率が向上したことが確認できる。

【0043】

【表1】

手法	学習データ			
	A	B	C	D
従来法	100.0	90.6	97.5	91.7
本発明 (個別VQ)	100.0	100.0	100.0	98.3

【0044】以上、本発明を実施例に基づいて具体的に説明したが、本発明は、前記実施例に限定されるものではなく、その要旨を逸脱しない範囲において種々変更し得ことはいうまでもない。

【0045】

【発明の効果】以上、説明したように、本発明によれば、事例からの学習によって動画像による動作認識系を構成することにより、従来の手法に比べ、自律的に対象や環境に適応でき、高度な認識が可能となるので、認識の成功率を向上（失敗の低減）することができる。

【0046】また、画像上でのモデルフィッティングを含まないで、実画像に対してもロバストな処理が実現できる。従って、本発明は、銀行や商店における不審動作監視、スポーツなどの動画から所望の動作部分の切り

【図面の簡単な説明】

【図1】 本発明の動作認識装置の一実施例の概略構成を示すブロック構成図、

10 【図2】 本実施例の機能構成を示すブロック構成図、

【図3】 本実施例の特徴抽出部で使したメッシュ特徴の説明図。

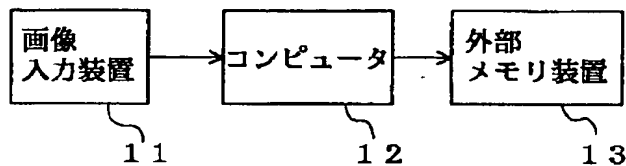
【符号の説明】

11…画像入力装置、12…コンピュータ、13…外部メモリ装置、21…画像入力部、22…画像用メモリ、23…特徴抽出部、24…特徴格納メモリ、25…量子化部、26…シンボル格納メモリ、27…尤度算出部、28…認識結果用メモリ、29…モデルパラメータ推定部、30…認識用状態遷移モデル格納メモリ、31…ベ

20

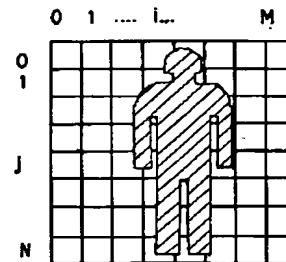
【図1】

図 1



【図3】

図 3

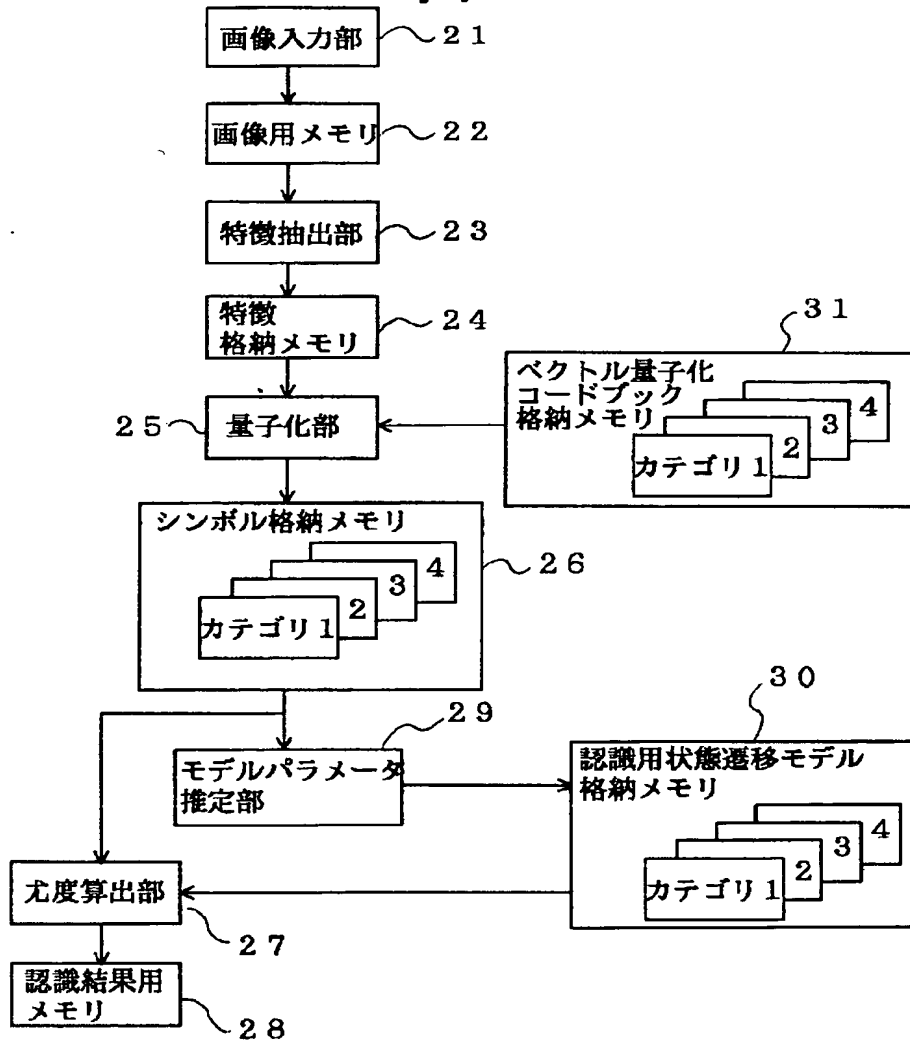


$$f=(a_{00},a_{01},\dots,a_{ij},\dots a_{Mv})$$

$$a_{ij}=\text{number of black mesh}(ij)/M_m N_m$$

【図2】

図 2



フロントページの続き

(72)発明者 石井 健一郎
 東京都千代田区内幸町一丁目1番6号 日
 本電信電話株式会社内